THE PREDICTION OF PILOT PERFORMANCE IN THE F-4 AIRCRAFT

Richard H. Shannon, et al

Naval Aerospace Medical Research Laboratory
Pensacola, Florida

9 July 1973

NAMRL-1186

THE PREDICTION OF PILOT PERFORMANCE IN THE F-4 AIRCRAFT

Lieutenant Richard H. Shannon, MSC, USN

and

Wayne L. Waag, Ph.D.

July 1973

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)*
Naval Aerospace Medical Research Laboratory
Naval Aerospace Medical Institute
Naval Aerospace & Regional Medical Center
Pensacola, Florida 32512

2a. REPORT SECURITY CLASSIFICATION
Unclassified

2b. GROUP

3. REPORT TITLE
THE PREDICTION OF PILOT PERFORMANCE IN THE F-4 AIRCRAFT

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

5. AUTHOR(S) *(First name, middle initial, last name)*
Richard H. Shannon, LT MSC USN
Wayne L. Wang

6. REPORT DATE
9 July 1973

7a. TOTAL NO. OF PAGES
~~8~~ 12

7b. NO. OF REFS
7

8a. CONTRACT OR GRANT NO.

b. PROJECT NO. MF51.524.002-5013DX5X

c.

d.

9a. ORIGINATOR'S REPORT NUMBER(S)
NAMRL-1186

9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*

10. DISTRIBUTION STATEMENT
Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY

13. ABSTRACT

In previous investigations, attempts were made to isolate the most critical skills and procedures within each stage of replacement air group (RAG) training in the F-4 aircraft. For each of the stages analyzed, a small set of items were selected on the basis that they could discriminate among replacement pilots according to their final RAG grade. On the basis of these isolated skills, two fleet evaluation questionnaires were developed to be used by operational F-4 squadron commanders. In addition to ratings on these two rating forms, squadron commanders were asked to report "critical incidents". These included such occurrences as accidents, incidents, and "tings-pulled". Data obtained from these two forms were used as the criterion measures in this investigation.

Selected test scores and flight grades from undergraduate pilot training were used as potential predictors. These were related to the criteria in a series of correlational and regression analyses. A number of significant relationships were obtained among the performance measures. Such results indicated the method used in developing the rating form to be a feasible one. Implications are discussed in terms of potential use for actual assignment of aviators to RAG training in the F-4 aircraft.

DD FORM 1 NOV 68 1473 (PAGE 1)

S/N 0101-807-6801

Unclassified

Security Classification

Unclassified
Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Criterion Development						
Aviation Personnel						
Aviation Training						
Pilot performance measurement						

DD FORM 1 NOV 68 1473 (BACK)
(PAGE 2)

Security Classification

1a

Approved for public release; distribution unlimited.

THE PREDICTION OF PILOT PERFORMANCE IN THE F-4 AIRCRAFT

Lieutenant Richard H. Shannon, MSC, USN

and

Wayne L. Waag, Ph.D

Bureau of Medicine and Surgery
MF51.524.002-5013DX5X

Approved by

Ashton Graybiel, M.D.
Assistant for Scientific Programs

Released by

Captain N. W. Allebach, MC, USN
Officer in Charge

9 July 1973

Naval Aerospace Medical Research Laboratory
Naval Aerospace Medical Institute
Naval Aerospace and Regional Medical Center
Pensacola, Florida 32512

i b

SUMMARY PAGE

THE PROBLEM

In previous investigations, attempts were made to isolate the most critical skills and procedures within each stage of replacement air group (RAG) training in the F-4 aircraft. For each of the stages analyzed, a small set of items were selected on the basis that they could discriminate among replacement pilots according to their final RAG grade. On the basis of these isolated skills, two fleet evaluation questionnaires were developed to be used by operational F-4 squadron commanders. In addition to ratings on these two rating forms, squadron commanders were asked to report "critical incidents". These included such occurrences as accidents, incidents, and "wings-pulled". Data obtained from these two forms were used as the criterion measures in this investigation.

FINDINGS

Selected test scores and flight grades from undergraduate pilot training were used as potential predictors. These were related to the criteria in a series of correlational and regression analyses. A number of significant relationships were obtained among the performance measures. Such results indicated the method used in developing the rating form to be a feasible one. Implications are discussed in terms of potential use for actual assignment of aviators to RAG training in the F-4 aircraft.

INTRODUCTION

The assessment of pilot performance entails two distinct problems, criterion definition and predictor selection. The definition of a criterion usually represents a compromise between some hypothetical "ideal" measure and measures which are readily available. One of the major obstacles to criterion specification lies in the diverse nature of the flight program. The progress of a student through the naval aviation training syllabus can be characterized by increased differentiation and specificity. It is only during the preflight and primary phases of training that students receive the same instruction. Following primary a student is assigned to either the jet, prop, or helo pipeline wherein the aircraft to be flown and syllabus requirements differ greatly. At each successive stage, comparability of training decreases and consequently the definition of a common criterion of pilot performance becomes increasingly difficult.

The postgraduate phase of training, or replacement air group (RAG) training, is even further diversified. Preparation of the pilot for actual fleet operations involves diverse aircraft with completely different handling characteristics and mission objectives. Furthermore, pilots entering RAG training usually differ in terms of previous experience. Some enter as nugget pilots directly from the training command; some have had one or more tours in other operational fleet aircraft, while others have had previous experience as an undergraduate training flight instructor. Still others enter directly from an interim shore billet in which they have received little flight time.

Once the aviator completes RAG training and is assigned to a fleet squadron, there is further diversification. Specific mission objectives and the operational environment are unlikely to be the same for different squadrons. Likewise, differences in experience level among squadron members will usually exist.

From this brief description of naval aviation, it is readily apparent that the development of adequate measures of pilot performance is greatly hindered by a lack of commonality within both training and operational commands. The fact remains, nonetheless, that the pilot is trained to become an integral part of a fleet squadron. Despite methodological and practical difficulties, the hypothetical "ideal" criterion must reflect the manner in which the pilot fulfills the mission objectives of his aircraft within the context of his squadron.

While fleet performance is the ultimate criterion, a number of intermediate criteria may be defined. Historically, most research efforts have been directed toward the prediction of intermediate criteria defined at the undergraduate level of training. The criterion most often adopted has been the success-failure dichotomy, or simply whether the student pilot successfully completes undergraduate training and receives his wings. The present Student Pilot Prediction System attests to the success of such efforts (1). Recently, the system has been extended to provide predictions of the final grade to aid in pipeline assignment.

At the RAG level, Bale, Rickus and Ambler (2) reported that certain grades obtained during undergraduate training were related to performance defined on a success/fail dichotomy. Specifically, combat-related skills appeared to be the best predictors. In a factor analytic study, Bale, Smith, and Ambler (3) found that certain clusters of skills could be extracted from stage

grades obtained during the undergraduate and postgraduate phases of training. Generally, those clusters appeared to be specific to the phase of training. From these data, a standardized RAG grading form was designed to be used by all communities and is at present awaiting implementation.

At the fleet level, Booth and Berkshire (4) reported a factor analysis of training command grades and fleet evaluations. Different factor structures were obtained from samples of jet pilots and helicopter pilots.

In assessing the previous literature, it is difficult to draw comparisons as a result of differences in sample group and time period. To date, an extensive investigation of a single community from training command to RAG to fleet has not been performed. In an attempt to provide such longitudinal continuity, a series of studies by Shannon and Waag (5, 6) have attempted to develop criterion measures of fleet performance for the F-4 community. In developing criteria, two approaches were taken. The first involved the application of the critical incident technique to fleet performance. An incident was said to occur if any one of a set number of operationally defined criterion events occurred. In this manner, a dichotomous criterion was established. Shannon and Waag (7) reported scores obtained during RAG training were significantly related to such a critical incident criterion.

The second approach focused upon the isolation of skills within RAG training as a result of their similarity to fleet operations. Using an item analytic procedure, the most important skills and procedures were isolated in which the final RAG grade served as the criterion. On the basis of these isolated skills, two fleet evaluation questionnaires were developed (6, 7).

The present study investigated the relationship between performance during undergraduate training and performance in the RAG and fleet. Specifically, an attempt was made to predict RAG performance as measured by the final RAG grade and fleet performance as estimated by: (1) critical incidents, and (2) ratings obtained from squadron commanders. In both cases, stage grades earned during undergraduate training served as a potential set of predictors.

METHOD

The sample group consisted of 173 replacement pilots assigned to RAG training in the F-4 aircraft between December 1969 and March 1972. All pilots were designated Category I, since they had never flown the F-4, and consequently were required to complete the entire RAG flight training syllabus. Of the sample, 120 completed training at VF-121, the West Coast Squadron, while the remaining 53 completed training at VF-101, the East Coast Squadron. For each pilot, the following information was obtained and used as a potential set of predictors: (1) Selection test scores, including the Aviation Qualification Test (AQT) and the Flight Aptitude Rating (FAR); (2) overall primary flight grade; (3) overall Basic flight grade; (4) stage grades obtained during the Advanced phase of undergraduate training; and (5) final RAG grade. Furthermore, each pilot was categorized according to experience level; that is, whether he entered RAG training directly from the undergraduate training command or had previous experience as a fleet pilot or as a training command instructor.

Rating forms developed in previous studies (6, 7) were sent to the respective fleet squadron command for pilots within the sample. For the East Coast sample, respective squadron commanders were asked to rate each pilot on two characteristics, Headwork and Basic Airwork (7). For the West Coast sample, a 17-item questionnaire was used (6). For each pilot, individual ratings on each item were summed to yield a total score. These total scores were standardized and transformed to T-scores for each coast in an attempt to equate the two rating forms. Squadron commanders were also asked to report any critical incidents such as accidents/incidents/mishaps attributable to human error, pulled-wings, turned-in-wings, etc. For each pilot, his final grade earned during RAG training was also obtained. These were also standardized for each coast in order to statistically control for possible differences. Regression analyses using a forward selection procedure were performed in an attempt to select those sets of variables which best predicted each of the criteria.

RESULTS

The matrix of intercorrelations from which the regression analyses were preformed is presented in Appendix A. Of the total sample, final RAG grades were obtained for 171 pilots. The results of the regression analyses predicting this criterion are presented in Table 1. It should be pointed out that the first variable to be entered, experience level, was initially forced into all of the analyses as a result of its moderator effects. Its importance will be discussed later.

Table 1

Summary of Regression Analysis Predicting Final RAG Grade

Variables Entered	Cumulative Multiple R
Experience Level	.314
Formation	.425
Transition	.458
Flight Aptitude Rating	.481
Basic Instruments	.490
Aviation Qualification Test	.500
Instrument Navigation	.501
Carrier Qualification	.513

From the potential set of predictors, the selection scheme produced a seven variable equation yielding a final multiple R of .513. Of all the variables, experience level was most related to the criterion suggesting that second tour pilots tend to receive better RAG grades. From Appendix A, it is apparent that experience level was also related to undergraduate training measures. This is not surprising since assignment following completion of undergraduate training is dependent upon his grades. In other words, those replacement pilots who enter RAG training directly from the training command are those receiving the highest grades. Yet those who enter from other fleet squadrons or instructor duty have had the benefit of added experience. Consequently, they tend to perform better in RAG training, despite the fact their grades during undergraduate training were lower. It is this fact which tends to reduce the magnitudes of the zero-order correlations of the undergraduate measures with the criterion. For this reason, experience-level was always the first variable to be entered in the regression analyses.

The two stage grades contributing the most non-redundant variance to the explanation of the criterion were the Advanced training stages of Formation and Transition. Interestingly enough, the two selection scores, the AQT and FAR, contained residual variance negatively related to the criterion.

Of the total sample, fleet ratings by squadron commanding officers were obtained for 99 pilots. The results of the regression analysis predicting this criterion are presented in Table 2. A five-variable equation yielding a multiple R of .401 was selected. Again the trend was toward second tour pilots receiving higher fleet evaluations. The final RAG grade as well as the Advanced formation and tactics grades were all positively associated with the criterion. Similar to the analysis predicting the final RAG grade, the Flight Aptitude Rating yielded a residual variance negatively associated with the criterion.

Table 2

Summary of Regression Analysis Predicting Fleet Evaluations

Variable Entered	Cumulative Multiple R
Experience Level	.150
Formation	.336
Tactics	.367
Final RAG Grade	.390
Flight Aptitude Rating	.401

Critical incident information was obtained for 102 pilots. Of these 25, or 24.5% of the total were credited with an incident. The results of the regression analysis predicting this criterion are presented in Table 3. A four-variable equation yielding a multiple R of .335 was selected. Of the variables entered into the equation, the final RAG grade showed most variance with the criterion.

Table 3

Summary of Regression Analysis Predicting Critical Incidents

Variable Entered	Cumulative Multiple R
Experience Level	.095
Final RAG Grade	.304
Transition	.319
Air-to-Ground	.335

DISCUSSION

The results indicate that, for each criterion measure, a certain proportion of the variance can be reliably explained from a subset of the potential predictors. It is interesting to note that somewhat different predictors emerged for each criterion. Of the total sample, both fleet

ratings and critical incident data were available for 96 pilots. The correlation between the criteria was found to be -.400 indicating a negative association between high ratings and the occurrence of an incident. While the sign of the correlation was in the expected direction, its magnitude seemed rather low. This suggests the possibility that fleet performance in the F-4 may be multi-dimensional in nature. In other words, adequate safe performance is not necessarily the same as adequate flight performance.

Consider the mission objectives of the F-4 aircraft. It was developed primarily for use as a tactical fighter, thereby demanding proficiency in combat-related skills. It seems likely that the fleet ratings best reflect such skills. Within the Advanced training phase, the Tactics stage grade and Formation grade were most highly correlated with the fleet ratings. This is not too surprising since the most common tactical configuration in the F-4 involves two aircraft flying in a section formation. In any case, combat-related skills appear to represent the most important components of the fleet ratings.

Several other interesting findings emerged. A trend was noted suggesting that pilots having prior flight experience as either a fleet pilot or training command instructor tended to receive higher fleet ratings. Furthermore, the FAR produced negative relationships with each of the criteria. However, the FAR was related to prior experience in that those having higher scores tended to become training command instructors. Consequently, those second-tour individuals having lower FARs tended to receive higher ratings as a possible result of their increased experience level.

In summary, the findings of this investigation suggest: (1) the final RAG grade can be reliably predicted from previous flight performance during undergraduate training ; (2) fleet ratings are most reflective of combat-related skills; (3) critical incidents are best predicted by performance in the RAG as estimated by the final RAG grade; and (4) pilots having previous experience tend to make better fleet pilots in the F-4 community.

REFERENCES

1. Aerospace Psychology Department, Manual for Use of the Student Pilot Prediction System. Naval Aerospace Medical Research Laboratory, Pensacola, Florida.

2. Bale, R. M., Rickus, G. M., and Ambler, R. K., Replacement air group performance as a criterion for naval aviation training. NAMRL-1126. Pensacola, Fla.: Naval Aerospace Medical Research Laboratory, 1970.

3. Bale, R. M., Smith, M. J., and Ambler, R. K., Factor analysis of undergraduate and postgraduate flight training grades. Aerospace Med., 43:372-375, 1972.

4. Booth, R. F., and Berkshire, J. R. Factor Analysis of Aviation Training Measures and Post-Training Performance Evaluations. NAMI 1050, Pensacola, Fla.: Naval Aerospace Medical Institute, 1968.

5. Shannon, R. H., Waag, W. L., and Ferguson, J. C. A new approach to criterion development in the replacement air group. NAMRL-1158. Pensacola, Fla.: Naval Aerospace Medical Research Laboratory, 1972.

6. Shannon, R. H., and Waag, W. L. Toward the development of a criterion for fleet effectiveness in the F-4 fighter community. NAMRL-1173, Pensacola, Fla.: Naval Aerospace Medical Research Laboratory, 1972.

7. Shannon, R. H., and Waag, W. L. Prediction of fleet success from performance on selected maneuvers in naval air training. Paper presented at the 80th annual meeting of the American Psychological Association, Honolulu, Hawaii. September, 1972.

APPENDIX A

Intercorrelation Matrix of the Predictor Variables Used in the Regression Analyses

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1.	Aviation Qualification Test	--	.309	-.342	-.112	-.120	-.096	.037	.087	-.066	-.039	.034	.095	.020	.099	.031	-.220	-.104	.060	-.146
2.	Flight Aptitude Rating		--	.340	.063	.075	-.022	.037	-.048	-.040	-.054	-.026	.110	.014	.067	.085	-.215	-.147	.063	-.169
3.	Primary Grade			--	.384	.271	.241	.375	.226	.255	.093	.199	.102	.266	.238	.222	-.091	-.299	.101	.015
4.	Basic Grade				--	.276	.300	.452	.152	.312	.183	.230	.212	.350	.228	.210	.084	-.182	.022	.052
5.	Transition					--	.319	.528	.195	.419	.209	.274	.334	.397	.402	.195	.076	-.462	.058	.044
6.	Basic Instruments						--	.455	.133	.384	.186	.146	.267	.267	.206	.164	-.106	-.390	.027	-.015
7.	Instrument Navigation							--	.316	.458	.283	.320	.424	.428	.360	.231	.042	-.443	-.052	.097
8.	Advanced Familiarization								--	.275	.108	.185	.214	.165	.195	.203	.016	-.188	-.035	-.027
9.	Formation									--	.300	.357	.426	.447	.434	.370	.149	-.373	.011	.223
10.	Night Familiarization										--	.138	.282	.309	.317	.191	.086	-.209	-.097	.015
11.	Operational Navigation											--	.272	.244	.333	.154	.011	-.201	.150	.129
12.	Air-to-Ground Weapons												--	.499	.382	.261	.004	-.339	-.078	.110
13.	Tactics													--	.426	.235	.052	-.399	-.060	.163
14.	Air-to-Air Weapons														--	.391	.075	-.283	-.038	.057
15.	Carrier Qualification															--	.076	-.215	-.085	.016
16.	Final RAG Grade																--	.314	-.304	.266
17.	Experience Level																	--	-.095	.150
18.	Critical Incident																		--	.402
19.	Fleet Evaluation																			--